

Scenario generation for portfolio selection problems with tail risk measure

Jamie Fairbrother^{*}, Amanda Turner^{*}, and Stein W. Wallace^{**}

^{*}STOR-i Centre for Doctoral Training, Lancaster University. United Kingdom

^{**}Department of Business and Management Science, Norwegian School of Economics.
Norway

November 17, 2015

Abstract

Tail risk measures such as the conditional value-at-risk are useful in the context of portfolio selection for quantifying potential losses in worst cases. However, for scenario-based problems these are problematic: because the value of a tail risk measure only depends on a small subset of the support of the distribution of asset returns, traditional scenario based methods, which spread scenarios evenly across the whole support of the distribution, yield very unstable solutions unless we use a very large number scenarios.

In this paper we propose a problem-driven scenario generation methodology for portfolio selection problems using a tail risk measure where the the asset returns have elliptical or near-elliptical distribution. Our approach in effect prioritizes the construction of scenarios in the areas of the distribution which correspond to the tail losses of feasible portfolios. The methodology is shown to work particularly well when the distribution of assets returns are positively correlated and heavy-tailed, and the performance is shown to improve as we tighten the constraints on feasible assets.

1 Introduction

In the portfolio selection problem one must decide how to invest in a collection of financial instruments with uncertain returns which in some way balances one's expected profit of the investment against its risk. In the typical set-up the uncertain returns are modeled by random variables, the total return of a portfolio is some linear combination of these, and riskiness is measured by a real-valued function of the total return which should in some way penalize potential large losses. This approach was first proposed by Markowitz [Mar52] who used variance as a risk measure.

The use of variance as a measure of risk is problematic for a few reasons. The foremost of these is perhaps that variance penalizes large profits as well as large losses. As a consequence, in the case where the returns of financial assets are not normally distributed, using the variance can lead to patently bad decisions; for instance, a portfolio can be chosen in favor of one which always has higher returns (see [You98] for an example of this). This particular issue can be overcome by using a “downside” risk measure, that is one which only depends on losses greater than the mean, or some other specified threshold, for example the semi-variance [Mar59, Chapter 9], mean regret [DR99], or value-at-risk [Jor96]. More recently, much research has been given to coherent risk measures, a concept introduced in [ADEH99]. These are risk measures which have sensible properties such as subadditivity, which in particular ensures that a risk measure incentivizes diversification of a portfolio. Using a coherent risk measure in a portfolio selection problem should avoid flawed decisions, such as the one cited in the case of variance.

In this work, we are interested in portfolio selection problems involving *tail risk measures*. These can be thought of as risk measures which only depend on the upper tail of a distribution above some specified quantile. A canonical example of a tail risk measure is the value-at-risk (VaR) [Jor96]. The

β -VaR is defined to be the β -quantile of a random variable. In portfolio selection problems this has the appealing interpretation as the amount of capital required to cover up to $\beta \times 100\%$ of potential losses. Thus, tail risk measures, in particular those which dominate the β -VaR, are useful as they can give us some idea of the amount capital at risk in the worst $(1 - \beta) \times 100\%$ of potential losses. Like variance, the value-at-risk is also problematic as it is not a coherent measure of risk. Specifically, it is not subadditive (see [Tas02] for example). Moreover, β -VaR leads to difficult and intractable problems when used in an optimization context. The conditional value-at-risk (CVaR), sometimes referred to as the expected shortfall, is another tail risk measure and can be roughly thought as the conditional expectation of a random variable above the β -VaR. It is both coherent [AT02], and more tractable in an optimization setting [RU00].

However, the use of risk measures, even coherent ones such as β -CVaR, is still problematic in portfolio selection problems where the asset returns are modeled with continuous probability distributions. This is because the evaluation of many risk measures for arbitrary continuously distributed returns would involve the evaluation of multidimensional integrals, and this becomes computationally infeasible when our problems involve many assets. On the other hand, the evaluation of such an integral reduces to a summation if the returns have a discrete distribution.

Scenario generation is the construction of a finite discrete distribution to be used in a stochastic optimization problem. This may involve fitting a parametric model to asset returns and then discretizing this distribution, or directly modeling them with a discrete distribution, for example via moment-matching [HKW03]. In either case, standard scenario generation methods struggle to adequately represent the uncertainty in problems using tail risk measures. This is because the value of a tail risk measure, by definition, only depends on a small subset of the support of a random variable, and typical scenario generation methods will spread their scenarios evenly across the whole support of the distribution. This means that the region on which the value of the tail risk depends, is represented by relatively few scenarios. Hence, unless there are a very large number of scenarios, the value of of tail risk measure is very unstable (see [KWVZ07] for example).

The natural remedy to this problem is to represent the regions of the distribution on which the tail risk measure depends with more scenarios. Intuition would tell us that these correspond to the “tails” of the distribution. However, for a multivariate distribution there is no canonical definition of the tails. If by tails, we simply mean the region where at least one of the components exceeds a large value, then the probability of this region quickly converges to one with the problem dimension, and thus prioritizing scenarios in this region will be of little benefit. Finding the relevant tails of the distribution is a non-trivial problem.

In the paper [FTW15] we addressed the problem of scenario generation for general stochastic programs using tail risk measures, and for this we defined the concept of a β -risk region. In portfolio selection, to each valid portfolio there is a distribution of losses (or returns). The β -risk region consists of all potential asset returns which lead to a loss in the β -tail for some portfolio. We have shown that the value of a tail risk measure in effect only depends on the distribution of returns in the risk region. Although characterizing this region in a convenient way is generally not possible, we have been able to do this for the portfolio selection problem when the asset returns are elliptically distributed. We have also proposed a sampling approach to scenario generation using these risk regions which prioritizes the generation of scenarios in the risk region. We demonstrated for simple examples that this methodology can produce scenario sets which yield better and more stable solutions than does basic sampling. In Sections 2 and 4 we review respectively the requisite theory of risk regions, and how risk regions can be used in scenario generation.

In this paper we address issues related to the application of this methodology to realistic portfolio selection problems. Firstly, we deal with how problem constraints are used to calculate the risk region. In [FTW15] we showed that for elliptically distributed returns, the risk region depends on the conic hull of the feasible region but we only did calculations for the case where this is the positive quadrant, that is, we only use the constraint of no short-selling. In practice, one may wish to impose other constraints on portfolios, such as quotas on the amount one can invest an asset or industry. In Section 3 we describe how the conic hull of the feasible region can be calculated from linear constraints, and how this is used to test whether or not a point lies in the risk region.

The effectiveness of the presented methodology depends directly on the probability of the risk region. In effect, the smaller the probability of the risk region, the more redundant scenarios we can discard. In [FTW15] we observed that the probability of this region tends to one as the problem dimension increases. In Section 5 we make some more general observations on how this probability varies with distribution, and in particular observe that for distributions with heavy tails and positive correlations, characteristics of typical stock return data, the probability of the risk region is low.

In practice it may not be appropriate to model asset returns with elliptical distributions as these are likely to exhibit non-elliptical features such as skewness [KW11]. Moreover, when the asset returns have an elliptical distribution, the portfolio selection problem may be solvable by more efficient methods [KBF07]. In Section 6 we test our methodology for a variety of distributions constructed from real data. We calculate the probability of the risk region for a range of constraints, and test the performance of our methodology for scenario generation and scenario reduction. We demonstrate here that when asset returns are near-elliptical in distribution, we can approximate its risk region with the risk region of an elliptical distribution to good effect.

Finally, in Section 7 we demonstrate for a difficult case study problem how our methodology can be applied. In particular we demonstrate how the addition of artificial constraints to the problem can be used to find better solutions.

2 Portfolio selection and risk regions

In this section we recall the requisite concepts and results from our previous paper [FTW15]. In particular we define the risk region for the portfolio selection problem and give a convenient characterization of this when asset returns have elliptical distributions.

2.1 Tail risk measures and risk regions

As mentioned above, a risk measure is a function of a real-valued random variable representing a loss. For $0 < \beta \leq 1$, a β -tail risk measure can be thought of as a function of a random variable which depends only on the upper $(1 - \beta)$ -tail of the distribution. The precise definition uses the *generalized inverse distribution function* or *quantile function*.

Definition 2.1 (Quantile function and β -tail risk measure). Suppose Z is a random variable with distribution function F_Z . Then the generalized inverse distribution function, or *quantile function* is defined as follows:

$$F_Z^{-1} : (0, 1] \rightarrow \mathbb{R} \cup \{\infty\}$$

$$\beta \mapsto \inf\{z \in \mathbb{R} : F_Z(x) \geq \beta\}$$

Now a β -tail risk measure is any function of a random variable, $\rho_\beta(Z)$, which depends only on the quantile function of a random variable above β .

Example 2.2 (Value at risk (VaR)). Let Z be a random variable, and $0 < \beta < 1$. Then, the β -VaR for Z is defined to be the β -quantile of Z :

$$\beta\text{-VaR}(Z) := F_Z^{-1}(\beta)$$

Example 2.3 (Conditional value at risk (CVaR)). Let Z be a random variable, and $0 < \beta < 1$. Then, the β -CVaR can be thought roughly as the conditional expectation of a random variable above its β -quantile. The following alternative characterization of β -CVaR [AT02] shows directly that it is a β -tail risk measure.

$$\beta\text{-CVaR}(Z) = \int_{\beta}^1 F_Z^{-1}(u) \, du$$

The observation that we exploit for this work is that very different random variables will have the same β -tail risk measure as long as their β -tails are the same. Such a situation is illustrated in Figure 1 for two discrete random variables.

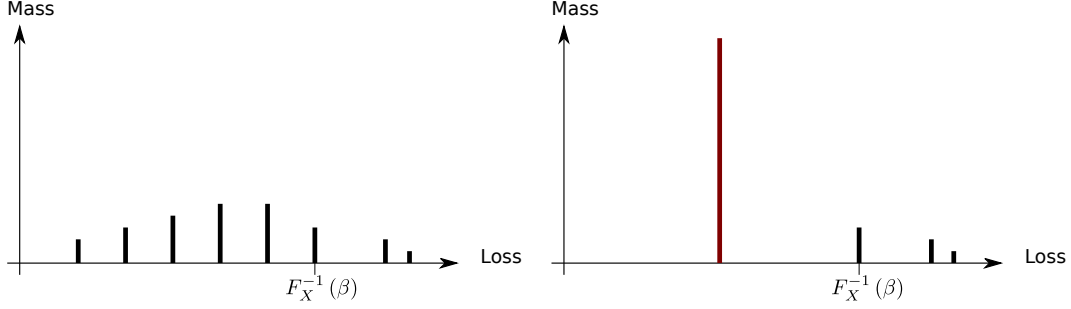


Figure 1: Two very different random variables with identical β -tails

In this paper we are interested in portfolio selection problems which use β -tail risk measures. We use the following basic set-up: we have a set of financial assets indexed by $i = 1, \dots, d$, by x_i we denote how much we invest in asset i , and by Y_i we denote the random future return of asset i . The profit associated to a particular investment decision $x = (x_1, \dots, x_d)$ and return $Y = (Y_1, \dots, Y_d)$ is $x^T Y = \sum_{i=1}^d x_i Y_i$. The loss associated to an investment decision is thus $-x^T Y$, and so for a given β -tail risk measure ρ_β we would like an investment with small risk $\rho_\beta(-x^T Y)$. The aim of a portfolio selection problem is to choose a decision which balances choosing a portfolio with high expected profit against choosing one with small risk. This typically corresponds to solving a problem of one of the following forms:

$$(i) \quad \underset{x \in \mathcal{X}}{\text{minimize}} \quad \rho_\beta(-x^T Y) \quad (P1)$$

$$\text{subject to } \mathbb{E}[x^T Y] \geq t,$$

$$(ii) \quad \underset{x \in \mathcal{X}}{\text{maximize}} \quad \mathbb{E}[x^T Y] \quad (P2)$$

$$\text{subject to } \rho_\beta(-x^T Y) \leq s,$$

$$(iii) \quad \underset{x \in \mathcal{X}}{\text{minimize}} \quad \lambda \rho_\beta(-x^T Y) + (1 - \lambda) \mathbb{E}[-x^T Y], \quad (P3)$$

where $0 \leq \lambda \leq 1$ and $\mathcal{X} \subset \mathbb{R}^d$ represents the set of valid portfolios. This feasibility region will typically encompass a constraint which specifies the amount of capital to be invested, and may include others which, for example the exclusion of short-selling, or a limit on the amount that can be invested in certain industries.

In [FTW15] we introduced the concept of a risk region for a stochastic program using a tail-risk measure. We define this now for the portfolio selection problem.

Definition 2.4 (Risk region). The β -risk region associated with the random vector Y and the feasible region $\mathcal{X} \subseteq \mathbb{R}^d$ is as follows:

$$\mathcal{R}_{Y, \mathcal{X}}(\beta) := \bigcup_{x \in \mathcal{X}} \{y \in \mathbb{R}^d : -x^T y \geq F_{-x^T Y}^{-1}(\beta)\}. \quad (1)$$

The risk region consists precisely of those outcomes of Y which have a loss in the β -tail of the loss distribution for *some* feasible portfolio. We refer to the complement of the risk region as the non-risk region and this consists of outcomes which never lead to a loss in the β -tail; it can be written as follows:

$$\mathcal{R}_{Y, \mathcal{X}}(\beta)^c = \bigcap_{x \in \mathcal{X}} \{y \in \mathbb{R}^d : -x^T y < F_{-x^T Y}^{-1}(\beta)\}. \quad (2)$$

Note that since this set is the intersection of half-spaces, it is convex.

For a discrete distribution of returns, we can easily calculate the β -quantile of the loss for a particular portfolio by calculating the loss for all scenarios and ordering them based on this. In Figure 2 we illustrate a scenario set of returns for two hypothetical assets, along with the line separating those scenarios with loss above and below the β -quantile for the portfolio $x = (\frac{1}{2}, \frac{1}{2})$.

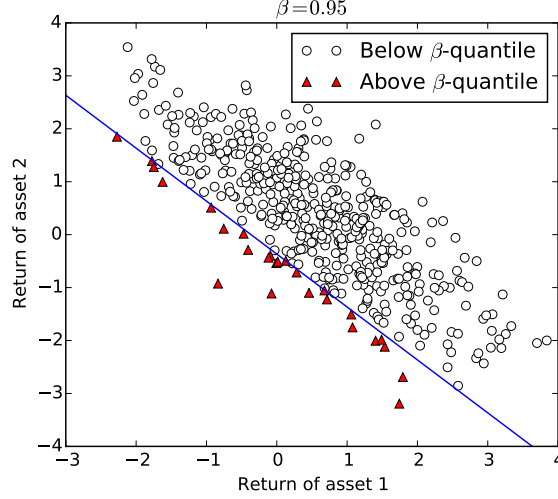


Figure 2: Scenarios with loss above and below β -quantile for one portfolio

The risk region is the union over all feasible portfolios of the half spaces of points with returns above the β -quantile. We can find this region by brute force, and this is illustrated on the left-hand side of Figure 3. Also illustrated in this figure is the set of returns where all the mass in the non-risk region has been aggregated into a single scenario. We call the latter the *aggregated scenario set*. As is also shown in the figure, the β -quantile lines do not change after aggregation and so neither does the corresponding value of any β -tail risk measure. Aggregating scenarios leads to more concise optimization problems which are easier to solve.

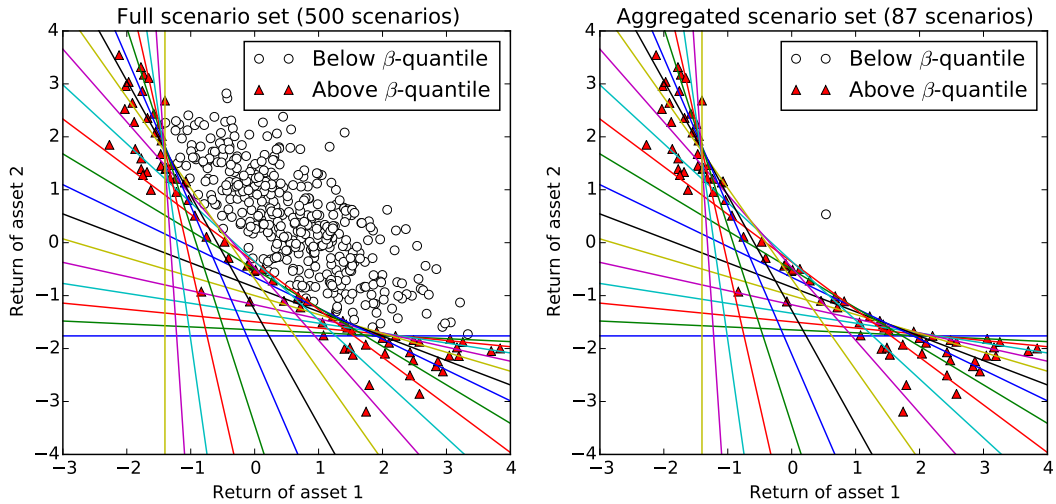


Figure 3: Return points with loss below the β -quantile for all non-negative portfolios (left) and aggregated scenario set (right)

The transformed random vector where all mass in a region has been concentrated into its conditional expectation plays a special role in this work. We call this the *aggregated random vector*.

Definition 2.5 (Aggregated Random Vector). For some set $\mathcal{R} \supseteq \mathcal{R}_{Y,\mathcal{X}}$ the *aggregated random vector* is

defined as follows:

$$\psi_{\mathcal{R}}(Y) := \begin{cases} Y & \text{if } Y \in \mathcal{R}, \\ \mathbb{E}[Y|Y \in \mathcal{R}^c] & \text{otherwise.} \end{cases}$$

In [FTW15] we showed that under mild conditions the value of a tail risk measure is completely determined by the the distribution of the random vector Y in the risk region. That is, the values of the tail risk measure of any two random vectors with identical distributions in the risk region will be the same for all feasible decisions.

Theorem 2.6. Let $\mathcal{R} \supseteq \mathcal{R}_{Y,\mathcal{X}}(\beta)$ be such that for all $x \in \mathcal{X}$ the following condition holds:

$$\mathbb{P}\left(Y \in \{y : z' < -x^T y \leq F_{-x^T Y}^{-1}(\beta)\} \cap \mathcal{R}\right) > 0 \quad \forall z' < F_{-x^T Y}^{-1}(\beta). \quad (3)$$

If \tilde{Y} is a random vector for which the following holds:

$$\mathbb{P}(Y \in \mathcal{A}) = \mathbb{P}(\tilde{Y} \in \mathcal{A}) \quad \text{for any } \mathcal{A} \subseteq \mathcal{R}, \quad (4)$$

then $\rho_{\beta}((f(x, Y))) = \rho_{\beta}(f(x, \tilde{Y}))$ for all $x \in \mathcal{X}$, for any β -tail risk measure ρ_{β} .

The technical condition (3) precludes certain degenerate cases. If \mathcal{R} is convex, we have that $\mathbb{E}[Y|Y \in \mathcal{R}^c] \in \mathcal{R}^c$ in which case the aggregated random vector defined above has, by definition, the same distribution as Y in the risk region. The aggregated random vector has the additional property of preserving the overall expected return of the original random vector. The following corollary taken from [FTW15] summarizes this result and provides sufficient conditions so that (3) holds.

Corollary 2.7. Suppose $\mathcal{R}_{Y,\mathcal{X}}(\beta) \subseteq \mathcal{R} \subset \mathbb{R}^d$, Y is a continuous random vector with support $\mathcal{Y} = \mathbb{R}^d$, and \mathcal{X} contains at least two linearly independent elements. Then Y satisfies (3). In addition, if \mathcal{R} is convex then $\tilde{Y} = \psi_{\mathcal{R}}(Y)$ satisfies condition (4) and so for all $x \in \mathcal{X}$ we have:

$$\begin{aligned} \rho_{\beta}(-x^T Y) &= \rho_{\beta}(-x^T \tilde{Y}), \\ \mathbb{E}[x^T Y] &= \mathbb{E}[x^T \tilde{Y}]. \end{aligned}$$

2.2 Risk regions for elliptical distributions

In order to exploit risk regions for scenario generation one has to be able to characterize these in a way which allows one to conveniently test whether or not a point belongs to it. In our previous paper, we were able to do this in the case where the asset returns have *elliptical distributions*. Elliptical distributions are a general class of distributions which include, among others, multivariate Normal and multivariate t -distributions. See [FKN89] for a full overview of the subject.

Definition 2.8 (Elliptical Distribution). Let $X = (X_1, \dots, X_d)$ be a random vector in \mathbb{R}^d , then X is said to be *spherical*, if

$$X \sim UX \quad \text{for all orthonormal matrices } U$$

where \sim means the two operands have the same distribution function.

Let Y be a random vector in \mathbb{R}^d , then Y is said to be *elliptical* if it can be written $Y = PX + \mu$ where $P \in \mathbb{R}^{d \times d}$ is non-singular, $\mu \in \mathbb{R}^d$, and X is random vector with spherical distribution. Such an elliptical distribution will be denoted Elliptical(X, μ, P).

This definition says that a random vector with a spherical distribution is rotation invariant, and that an elliptical distribution is an affine transformation of a spherical distribution. Elliptical distributions are convenient in the context of portfolio selection as we can write down exactly the distribution of loss of a portfolio:

$$-x^T Y \sim \|Px\| X_1 - x^T \mu,$$

and so, the β -quantile of the loss $-x^T Y$ is as follows:

$$F_{-x^T Y}^{-1}(\beta) = \|Px\| F_{X_1}^{-1}(\beta) - x^T \mu.$$

For elliptically distributed returns, we can thus rewrite the risk region in (1) as follows:

$$\mathcal{R}_{Y, \mathcal{X}}(\beta) := \bigcup_{x \in \mathcal{X}} \{y \in \mathbb{R}^d : -x^T y \geq \|Px\| F_{X_1}^{-1}(\beta) - x^T \mu\}. \quad (5)$$

In this form it is still difficult to check whether a given point $\tilde{y} \in \mathbb{R}^d$ belongs to it. In [FTW15] we provided a more convenient characterization of the risk region for elliptical returns. This characterization makes use of the conic hull of the feasible region $\mathcal{X} \subset \mathbb{R}^d$.

Definition 2.9 (Convex cones and conic hull). A set $K \subset \mathbb{R}^d$ is a cone if for all $x \in K$ and $\lambda \geq 0$ we have $\lambda x \in K$. A cone is convex if for all $x_1, x_2 \in K$ and $\lambda_1, \lambda_2 \geq 0$ we have $\lambda_1 x_1 + \lambda_2 x_2 \in K$. The conic hull of a set $\mathcal{A} \subset \mathbb{R}^d$ is the smallest convex cone containing \mathcal{A} , and is denoted $\text{conic}(\mathcal{A})$.

For example, suppose that our feasible region consists of portfolios with non-negative investments (i.e. no short-selling) and whose total investment is normalized to one, that is:

$$\mathcal{X} = \{x \in \mathbb{R}^d : \sum_{i=1}^d x_i = 1, \quad x_i \geq 0 \text{ for each } i = 1, \dots, d\},$$

then the conic hull of this is the positive quadrant, that is $\text{conic}(\mathcal{X}) = \mathbb{R}_+^d$. The alternative characterization also makes use of projections.

Definition 2.10 (Projection). Let $C \subset \mathbb{R}^d$ be a closed convex set, then for any point $y \in \mathbb{R}^d$ we define its projection onto C to be the unique point $p_C(y) \in C$ such that

$$\inf_{x \in C} \|x - y\| = \|p_C(y) - y\|.$$

We are now ready to give a characterization of our risk region. The following result was proved in [FTW15].

Theorem 2.11. Suppose $Y \sim \text{Elliptical}(X, P, \mu)$, $\mathcal{X} \subseteq \mathbb{R}^d$ is convex and let $K = \text{conic}(\mathcal{X})$. Then the risk region can be characterized exactly as follows:

$$\mathcal{R}_{Y, \mathcal{X}}(\beta) = P^T \left(\{\tilde{y} \in \mathbb{R}^d : \|p_{K'}(\tilde{y} - \mu)\| \geq F_{X_1}^{-1}(\beta)\} \right), \quad (6)$$

where $K' = PK$ is a linear transformation of the conic hull K .

When $K = \mathbb{R}^d$ the complement of the region in (6) has a convenient geometric description; it is an open ellipsoid.

3 Projections and the conic hull

The characterization of the risk region for portfolio selection problems given in (6) relies on one being able to calculate the conic hull of the set of feasible portfolios, and also the ability to project points onto a transformation of this. In Section 3.1 we show how one can find the conic hull of the feasible region for typical constraints of a portfolio selection problem. This conic hull is a *finitely generated cone*. In 3.2 we show how one can project points onto this type of cone.

3.1 Conic hull of feasible region

In portfolio problems, the feasible region is usually defined by linear constraints, that is $\mathcal{X} = \{x \in \mathbb{R}^d : Ax \leq b\}$, where $A \in \mathbb{R}^{m \times d}$ and $b \in \mathbb{R}^m$. That is, the feasible region is the intersection of a finite number

of half-spaces. It is a well-known fact that any such intersection can be written as a the convex hull of a finite number of points plus the conical combination of some more points (see Theorem 1.2 in [Zie08] for example). That is, there exists $x_1, \dots, x_k \in \mathbb{R}^d$ and $y_1, \dots, y_l \in \mathbb{R}^d$ such that

$$\mathcal{X} = \left\{ \sum_{i=1}^k \lambda_i x_i + \sum_{j=1}^l \nu_j y_j : \lambda, \nu \geq 0, \sum_{i=1}^k \lambda_i = 1 \right\}. \quad (7)$$

The conic hull of this region is the following *finitely generated cone*:

$$\text{conic}(\mathcal{X}) = \left\{ \sum_{i=1}^k \lambda_i x_i + \sum_{j=1}^l \nu_j y_j : \lambda, \nu \geq 0 \right\}.$$

To express the intersection of half-spaces in the form (7), we could use *Chernikova's algorithm* (also known as the double description method) [Che65], [LV92]. Every finitely generated cone can also be written as a *polyhedral cone*, that is, of the form $\{x \in \mathbb{R}^d : Dx \geq 0\}$, and vice versa (see [Zie08, Chapter 1]). Chernikova's algorithm again provides a concrete method for going between these two different representations. Although these two representations are mathematically equivalent, as we shall see, they are algorithmically different.

We will suppose the constraints for our portfolio selection problem have the following form:

$$\mathcal{X} = \left\{ \begin{array}{l} \mathbf{1}^T x = c \\ x \in \mathbb{R}^d : a_i^T x \leq b_i \quad \text{for } i = 1, \dots, m, \\ x \geq 0, \end{array} \right\} \quad (8)$$

where $\mathbf{1}$ is column vector of ones and $c > 0$. The first of these constraints specifies the total of amount of capital to be invested, the inequalities represent other constraints such as quotas on the amount one can invest in a specific company or industry. In this case, we can describe immediately the conic hull as a polyhedral cone.

Proposition 3.1. Let \mathcal{X} be the set defined in (8) and let

$$\mathcal{Y} = \left\{ x \in \mathbb{R}^n : \left(\frac{b_i}{c} \mathbf{1} - a_i \right)^T x \geq 0 \text{ for } i = 1, \dots, m, x \geq 0 \right\}$$

then $\text{conic}(\mathcal{X}) = \mathcal{Y}$.

Proof. Given that \mathcal{X} is convex, to show that $\text{conic}(\mathcal{X}) = \mathcal{Y}$, it suffices to show that

$$x \in \mathcal{Y} \setminus \{0\} \iff \exists \lambda > 0 \text{ such that } \lambda x \in \mathcal{X}.$$

We demonstrate first the forward implication. Suppose $x \in \mathcal{Y} \setminus \{0\}$. Then, given that $x > 0$, we must have $v := \mathbf{1}^T x > 0$. Then, setting $\lambda = \frac{c}{v}$, we have

$$\mathbf{1}^T(\lambda x) = v \frac{c}{v} = c.$$

Since \mathcal{Y} is a cone, we have $\lambda x \in \mathcal{Y}$, hence

$$\begin{aligned} & \left(\frac{b_i}{c} \mathbf{1} - a_i \right)^T \frac{c}{v} x \geq 0 \\ \therefore & \frac{c}{v} a_i^T x \leq \frac{b_i}{c} \frac{c}{v} \underbrace{\mathbf{1}^T x}_{=v} \\ \therefore & a_i^T \left(\frac{c}{v} x \right) \leq b_i \end{aligned}$$

and so $\lambda x \in \text{conic}(\mathcal{X})$.

We now prove the backwards implication. Suppose $x \in \text{conic}(\mathcal{X}) \setminus \{0\}$. Then there exists $\lambda > 0$ such that $\lambda x \in \mathcal{X}$, that is

$$\begin{aligned} \mathbf{1}^T \lambda x &= c \\ a_i^T \lambda x &\leq b_i \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{a_i^T \lambda x}{\mathbf{1}^T \lambda x} &\leq \frac{b_i}{c} \\ \text{and so } \left(\frac{b_i}{c} \mathbf{1} - a_i \right)^T x &\geq 0. \end{aligned}$$

Hence $x \in \mathcal{Y}$ as required. \square

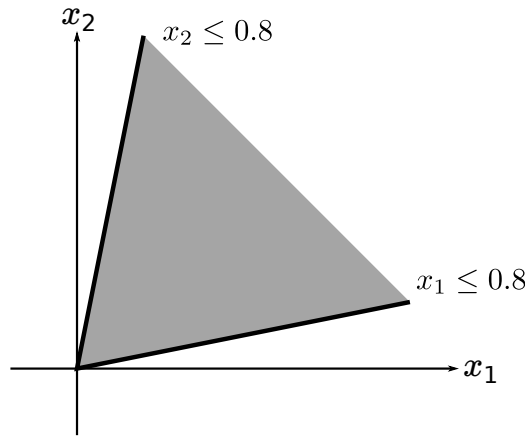


Figure 4: Conic hull from simple quota constraints given $x_1 + x_2 = 1$ and $x_1, x_2 \geq 0$

Figure 4 shows how simple constraints in \mathbb{R}^2 affect the conic hull of the feasible region given the total investment and positivity constraints.

3.2 Projection onto a finitely generated cone

First, suppose that we can represent the conic hull of the feasible region $\mathcal{X} \subset \mathbb{R}^d$ as a finitely generated cone $K = \{Ay : y \geq 0\}$ where $A \in \mathbb{R}^{k \times d}$. By definition, the projection of a point $x_0 \in \mathbb{R}^d$ can be found by solving the following quadratic program:

$$\underset{y \geq 0}{\text{minimize}} \quad \|Ay - x_0\|_2^2 \tag{9}$$

In particular, if y^* is the optimal solution then $p_K(x_0) = Ay^*$. By formulating the KKT conditions [BV04, Chapter 5] of this problem, it can be seen that this problem is equivalent to solving the following linear complementarity problem (LCP):

$$\begin{aligned} \text{Find } y, z \in \mathbb{R}^d \text{ such that} \\ z - A^T Ay &= -A^T x_0 \\ z^T x &= 0 \\ x, z &\geq 0. \end{aligned}$$

LCPs can be solved by more specialized algorithms than standard quadratic programs such as Lemke's algorithm [CPS92].

Now, suppose instead we have a polyhedral characterization of the conic hull, that is a cone of the form:

$$K = \{x \in \mathbb{R}^d : Bx \geq 0\}. \quad (10)$$

The projection of a point $x_0 \in \mathbb{R}^d$ onto the polyhedral cone in (10) is the solution of the following quadratic program:

$$\begin{aligned} & \underset{x}{\text{minimize}} \quad \|x - x_0\|_2^2 \\ & \text{subject to} \quad Bx \geq 0. \end{aligned}$$

Although the former problem in (9) can be solved using specialized algorithms, we will in practice use both approaches. For conic hulls with a small number of extremal rays, for example $K = \mathbb{R}_+^d$ we will use the former method. As we add more constraints to the problem, we have found from experience that the number of extremal rays can exponentially increase, which for the former approach leads to cumbersome large LCP problems. In this case we will use the polyhedral representation for projection.

4 Scenario generation

In this Section we present how risk regions can be exploited for the purposes of scenario generation. In Section 4.1 we propose two specific methods which work essentially by prioritizing the construction of scenarios in the risk region. These methods are specifically adapted to asset returns which have elliptical distributions, and so in Section 4.2 we discuss their usage for non-elliptic distributions. Finally, in Section 4.3 we discuss how the effectiveness of our methodology can be improved through the addition of artificial constraints to our problem.

4.1 Aggregation sampling and reduction

In this section we will assume that asset returns have elliptical distributions from which we can sample. In [FTW15] we proposed two methods to exploit risk regions. The first of these allows the user to specify the final number of scenarios in advance. The algorithm, which we call *aggregation sampling*, samples scenarios, aggregating all samples in the non-risk region and keeping all in the risk region, until we have the required number of risk scenarios. This is described in Algorithm 1.

Let $q = \mathbb{P}(Y \in \mathcal{R}_{Y,\mathcal{X}}^c)$ be the probability of the non-risk region, and n be the number of risk scenarios required. Define $N(n)$ to be the effective sample size from aggregation sampling, that is, the number of draws until the algorithm terminates¹. The quantity $N(n)$ is a random variable:

$$N(n) \sim n + \mathcal{NB}(n, q),$$

where $\mathcal{NB}(N, q)$ denotes a *negative binomial* random variable. Recall that a negative binomial random variable $\mathcal{NB}(n, q)$ is the number of failures in a sequence of Bernoulli trials with probability of success q until n successes have occurred. The expected effective sample size of aggregation sampling is thus as follows:

$$\mathbb{E}[N(n)] = n + n \frac{q}{1-q}$$

The expected effective sample size can be thought of as the sample size required for basic sampling to produce the same number of scenarios in the risk region. The difference between the desired number of risk scenarios, and expected effective sample size is proportional to the ratio $\frac{q}{1-q}$. In particular, as the probability of the non-risk region approaches one, this gain tends to infinity.

The converse to aggregation sampling is sampling a set of a given size n and then aggregating all scenarios in the risk region of the underlying distribution. We call this *aggregation reduction*. This can be viewed as a sequence of n Bernoulli trials, where success and failure are defined in the same way as

¹For simplicity of exposition we discount the event that the while-loop of the algorithm terminates with $n_{\mathcal{R}^c} = 0$ which occurs with probability q^n

```

input :  $N_{\mathcal{R}}$  number of required risk scenarios,  $\beta$  level of tail risk measure,  $K$  conic hull of
feasible region,  $(X, P, \mu)$  elliptical distribution
output:  $\{(y_s, p_s)\}_{s=1}^{N_{\mathcal{R}}+1}$  scenario set
 $n_{\mathcal{R}^c} \leftarrow 0$ ,  $n_{\mathcal{R}} \leftarrow 0$ ,  $y_{\mathcal{R}^c} \leftarrow 0$ ;
 $K' \leftarrow PK$ ;
while  $n_{\mathcal{R}} < N_{\mathcal{R}}$  do
    Sample new point  $y$ ;
     $y_{\text{trans}} \leftarrow P^{-T}(y - \mu)$ ;
    if  $\|p_{K'}(y_{\text{trans}})\| \leq F_{X_1}^{-1}(\beta)$  then
         $y_{\mathcal{R}^c} \leftarrow \frac{1}{n_{\mathcal{R}^c}+1}(n_{\mathcal{R}^c}y_{\mathcal{R}^c} + y)$ ;
         $n_{\mathcal{R}^c} \leftarrow n_{\mathcal{R}^c} + 1$ ;
    end
    else
         $y_{n_{\mathcal{R}}} \leftarrow y$ ;
         $n_{\mathcal{R}} \leftarrow n_{\mathcal{R}} + 1$ ;
    end
end
foreach  $i$  in  $1, \dots, N_{\mathcal{R}}$  do  $p_i \leftarrow \frac{1}{n_{\mathcal{R}^c} + n_{\mathcal{R}}}$ ;
if  $n_{\mathcal{R}^c}^e \geq 1$  then
     $y_{N_{\mathcal{R}}+1} \leftarrow y_{\mathcal{R}^c}$ ,  $p_{N_{\mathcal{R}}+1} \leftarrow \frac{n_{\mathcal{R}^c}}{n_{\mathcal{R}^c} + n_{\mathcal{R}}}$ ;
end
else
    Sample new point  $y$ ;
     $y_{N_{\mathcal{R}}+1} \leftarrow y$ ,  $p_{N_{\mathcal{R}}+1} \leftarrow \frac{1}{N_{\mathcal{R}}+1}$ ;
end

```

Algorithm 1: Aggregation sampling algorithm for risk region of an elliptical distribution

described above. The number of scenarios in the reduced sample, $R(n)$ is as follows:

$$R(n) \sim n - \mathcal{B}(n, q) + 1$$

where $\mathcal{B}(n, q)$ denotes a binomial random variable. The expected reduction in scenarios in aggregation reduction is thus $nq - 1$.

The reason that aggregation sampling and aggregation reduction work is that, for large samples, they are equivalent to sampling from the aggregated random vector, and Corollary 2.7 tells us that this random vector has the correct tail risk measure and expectation. See [FTW15, Section 4] for a full proof of the consistency of these algorithms.

In the above methods, for every sampled point we must be able to test whether the magnitude of its projection onto a cone is above or below a given threshold. As explained in Section 3.2, the projection of a point onto a finite cone involves solving a small LCP or quadratic program and so for large sample sizes and high dimensions this will become computationally expensive. However, given that each sample is independent of every other, this algorithm is naturally parallelizable. It may also be possible to make the algorithm more efficient if for any point $y \in \mathbb{R}^d$, a way could be found of testing the condition $\|p_K(y)\| \leq \alpha$ directly without calculating the full projection $p_K(y)$. For example, the quadratic program used to calculate the projection could be solved only to an accuracy sufficient to test this condition.

4.2 Approximation of risk regions

The above algorithm is specifically adapted to risk regions of elliptically distributed returns. However, the utility of using our scenario generation methodology with only elliptical distributions is limited. Firstly, it may be unrealistic to model returns with elliptical distributions. Real financial returns may exhibit properties which elliptical distributions do not, such as skewness. Secondly, using elliptical distributions may allow one to formulate the optimization problem in a more convenient way. For example, when

using the β -CVaR as a tail-risk measure, we have the following relation:

$$\beta\text{-CVaR}(-x^T Y) = -\|Px\| \beta\text{-CVaR}(X_1) - x^T \mu$$

which may mean we can solve the optimization problems with interior point or quadratic programming algorithms. See [KBF07] for more details.

In this work, we put forward the idea that risk regions of elliptical distributions can be used for aggregation sampling and reduction, for distributions which are near-elliptical. We propose to use the risk region of an elliptical distribution as a surrogate for the risk region of a near-elliptical distribution.

The danger of approximating the risk region is that if for a particular decision, the β -quantile, is attained inside the surrogate non-risk region (that is the surrogate risk region is too small), then the value of the tail-risk measure may be distorted. On the other hand, if the surrogate risk region contains the true risk region (that is, it contains the true risk region) then Theorem Corollary 2.7 guarantees that the associated aggregated random vector has the correct tail risk measure. However, we should be cautious about constructing a surrogate risk region which is excessively large. If this is the case then the probability of the risk region may be very large, which means there will be little benefit in aggregation.

Through a careful probabilistic analysis of the distribution of returns for valid portfolios, one may be able to construct a surrogate risk region which tightly covers the true risk region. If this is not possible, one way to mitigate against the danger of using a surrogate risk region which is too small would be to represent the non-risk region with several points rather than a single point. For example, instead of aggregating all sampled points in the non-risk region, one could apply a clustering algorithm to these such as k -means. For simplicity, we will only test the basic aggregation methods which represents the non-risk region with a single point. For the non-elliptical distributions used in this paper we are able to rely on heuristic rules to construct our surrogates.

The first class of non-elliptical distributions we use in this paper are known as multivariate Skew-t distributions [AC03]. This class of distributions generalizes the elliptical multivariate t-distributions through the inclusion of an extra set of parameters which regulate the skewness. In this case we approximate the risk region with the risk region of the corresponding t-distribution.

The second class we use are discrete distributions constructed using the moment matching algorithm of [HKW03]. These distributions have been applied previously to financial problems [KWVZ07]. This algorithm constructs scenario sets with a specified correlation matrix and whose marginals have specified first four moments. This algorithm works by first taking a sample from a multivariate Normal distribution, and then iteratively applying transformations to this until the difference between its marginal moments and correlation matrix are sufficiently close to their target values. Since the algorithm is initialized with a sample from an elliptical distribution, the final distribution is near elliptical and we approximate the risk region for these distributions with the risk region of a multivariate normal distribution with the same mean and covariance structure.

4.3 Ghost constraints

We noted above that the performance of our methodology improves as the probability of the non-risk region decreases. In particular, the expected effective sample size in aggregation sampling increases as the probability of the non-risk risk region increases. Now, by its definition (2) the non-risk regions shrinks as the problem becomes more constrained. This suggests that it may be helpful to add constraints to our problem which shrink the set of feasible portfolios, but which are not themselves active, in the sense that their presence does not affect the set of optimal solutions. We will refer to a constraint added to a problem to boost the performance of our methodology, loosely, as a *ghost constraint*.

Finding non-active constraints to add to our problem is non-trivial as it relies on some knowledge of the optimal solution set. We will resort to heuristic rules to choose ghost constraints. For example, one could constrain our set of feasible portfolios to some neighborhood of a good quality solution.

Verifying whether or not a ghost constraint is active is difficult in general for stochastic programs. For a deterministic objective function which is convex and for which all constraints are convex (and the optimal solution is unique) a constraint $\{x : g(x) \leq 0\}$ is non-active if and only if it is binding at the

optimal solution x^* , that is $g(x^*) = 0$. For a stochastic program, we are typically solving a scenario-based approximation and so a constraint which is not binding with respect to the scenario-based approximation may be binding with respect to the true problem and vice versa.

A rigorous test of whether a ghost constraint is active in the sense above is beyond the scope of this paper. We simply promote the idea here that ghost constraints may be a useful way of finding better solutions. We demonstrate the usage of ghost constraints on a difficult problem in Section 7.

5 Probability of the non-risk region

The benefit of aggregation sampling and reduction depends on the probability of the non-risk region. As was observed in [FTW15] the probability of the non-risk region tends to decrease as the problem dimension increases, but increases as we tighten our problem constraints, and as we increase β , the level of the tail risk measure. In this section we make some empirical observations on how this probability varies with heaviness of the tails, and the correlations of the distribution.

The first observation is that in the presence of positivity constraints, the probability of the risk region improves as the correlation between random variables increases. This can be seen in Figure 5 which plots the probability of the risk region as a function of correlation for some two-dimensional distributions. An intuitive explanation for this type of behavior is that in the case of positive correlations there is much more overlap in the risk regions of the individual portfolios.

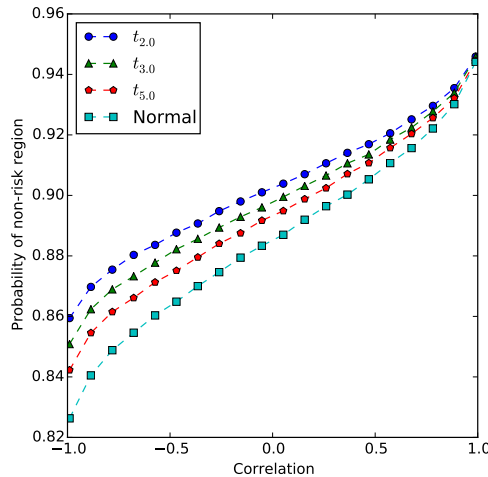


Figure 5: Correlation vs. Probability of non-risk region for 2-dimensional Normal distribution, positivity constraints and $\beta = 0.95$

The extent to which probabilities vary with correlation seems to be much greater in higher dimensions. In Figure 6 we have plotted for Normal returns and a range of dimensions, the probabilities of the non-risk region for a particular type of correlation matrix: $\Lambda(\rho) \in \mathbb{R}^{d \times d}$ where $\Lambda(\rho)_{ij} = \rho$ for $i \neq j$ and $\rho > 0$. In the case of $\rho = 0$, the probability decays very quickly to zero as the dimension increases, whereas as when ρ is close to one, the probability of the non-risk region approaches β for all dimensions.

Our next observation is that the probability of the non-risk region seems to increase as the tails of the distributions become heavier. In Figure 7 are plotted the probabilities of risk regions for some spherical distributions and a range of dimensions. Note that multivariate t-distributions have heavier tails than the multivariate Normal distribution, but the tails get lighter as the degrees of freedom parameter increases. This phenomenon can also be observed in Figure 5.

The observations made in this section suggest that the application of our methodology will be particularly effective when applied to real stock data tend to be positively correlated and have heavy tails.

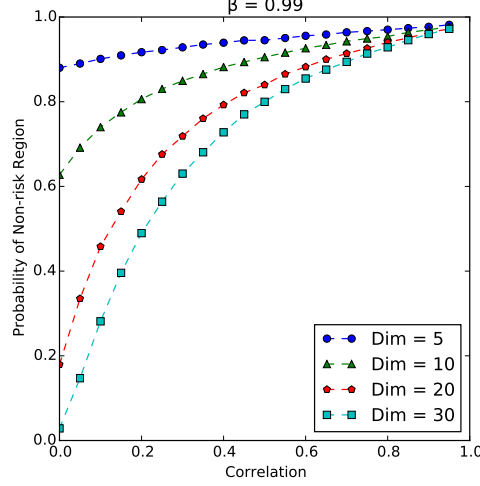


Figure 6: Probability of non-risk region for a range of correlation matrices and dimensions for Normal returns

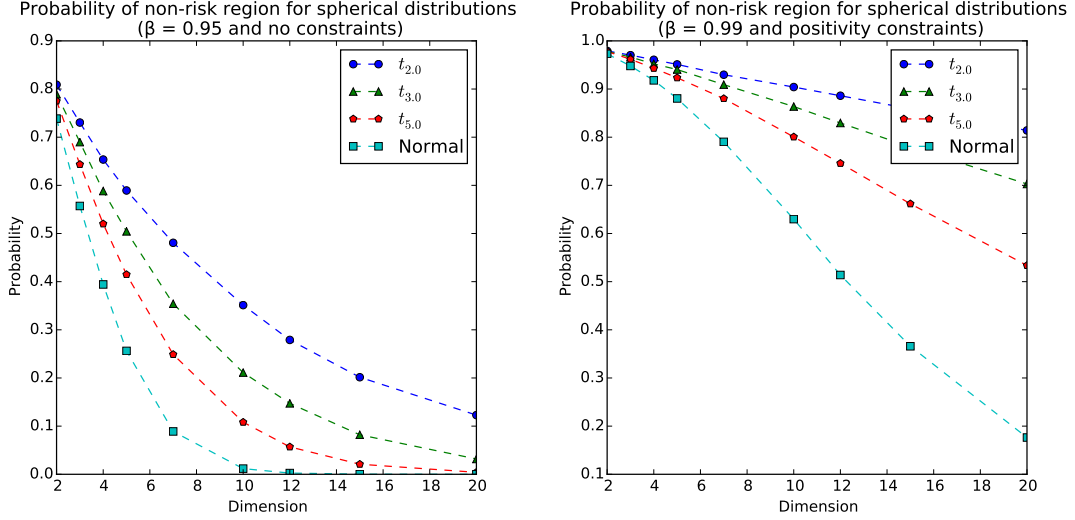


Figure 7: Probability of non-risk regions for different spherical distributions and dimensions

6 Numerical tests

In this Section we test the numerical performance of our methodology for realistic distributions. There are three parts to these tests: the calculation of the probability of the non-risk region for a range of distributions and constraints, the performance of aggregation sampling, and the performance of aggregation reduction. In Section 6.1 we describe our experimental set-up, in particular we justify the distributions constructed for these experiments. The remaining three sections detail the individual experiments and their results.

6.1 Experimental set-up

For robustness we will use several randomly constructed distributions for each family of distribution and each dimension we are testing. We construct these by fitting parametric distributions to real data. We use real data rather than randomly generating problem parameters for two reasons. Firstly, generating parameters which correspond to well-defined distributions can be problematic. For example, for the moment matching algorithm, there may not exist a distribution which has a given set of target moments

(see [KME00] and [JR51] for instance). Secondly, as was observed in Section 5, the probability of the non-risk region can vary wildly, and so it is most meaningful to test the performance of our methodology for distributions which are realistic for portfolio selection problems.

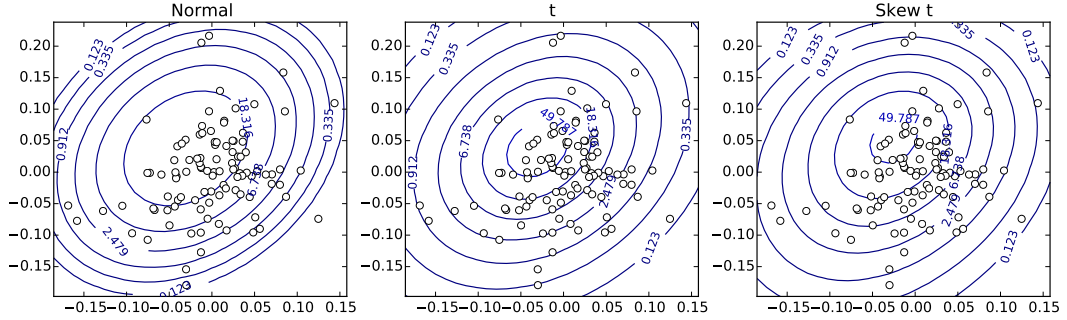


Figure 8: Contour plots of distributions fitted to financial return data for two assets

We construct our distributions from monthly return data from between January 2007 and February 2015 for 90 companies in the FTSE 100 index. For each dimension in the test, we randomly sample five sets of companies of that size, and for each of these sets fit Normal, t distributions and Skew- t distributions to the associated return data. Figure 8 shows for two stocks the contours of the fitted density functions overlaying the historical return data. For the t distributions we fix the degrees of freedom parameter to 4.0. This is so that we can more easily compare the effect of heavier tails on the results of our tests. We allow the corresponding parameter for the skew- t distributions to be fitted from the data.

These three distributions are fitted to the data through maximum likelihood estimation, weighing each observation equally; our aim here is not to construct distributions which accurately capture the uncertainty of future returns, but to simply construct distributions which are realistic for this type of problem. We also use scenario sets constructed using the moment-matching algorithm. For each random set of companies, we calculate all the required marginal moments and correlations from their historical returns, and use these as input to the moment-matching algorithm. To allow us to compare results, the same constructed distributions are used across the three sets of numerical tests.

Throughout this section we use the β -CVaR as our tail risk measure. This is not only because the β -CVaR leads to tractable scenario-based optimization problems, but also for elliptically distributed returns we can evaluate the β -CVaR exactly which provides us with a means to evaluate the true performance of the solutions yielded by the approximate scenario-based problems. In addition, to ensure that the non-risk region does not have negligible probability, we will assume that we always have positivity constraints on our investments (i.e. no short selling).

6.2 Probability of non-risk region with quota constraints

We first estimate the probability of the non-risk region for a range of distributions, dimensions and constraints. We calculate these probabilities only for the Normal and t distributions as skew- t distributions and moment matching scenario sets use surrogate risk regions based on these distributions. The main purpose of this is to provide intuition about under what circumstances the methodology is effective: there is little to be gained from aggregating scenarios in a non-risk region of negligible probability.

For each distribution we sample 2000 scenarios and calculate the proportion of points in the non-risk region for different levels of β and constraints. In particular, for each dimension we calculate for $\beta = 0.95$, and $\beta = 0.99$, and for a range of *quotas*. The feasible region corresponding to quota $0 < q < 1$ is $\{x \in \mathbb{R}^d : 0 \leq x_i \leq q \text{ for } i = 1, \dots, d, \sum_{i=1}^d x_i = 1\}$. Quotas are quite a natural constraint to use in the portfolio selection problem as they ensure that a portfolio is not overexposed to one asset. The quotas may also be viewed as ghost constraints to be used in cases where the probability of the non-risk region with only positivity constraints is too small.

In Figure 9 for each each dimension we have tested we have plotted the results of one trial. The full

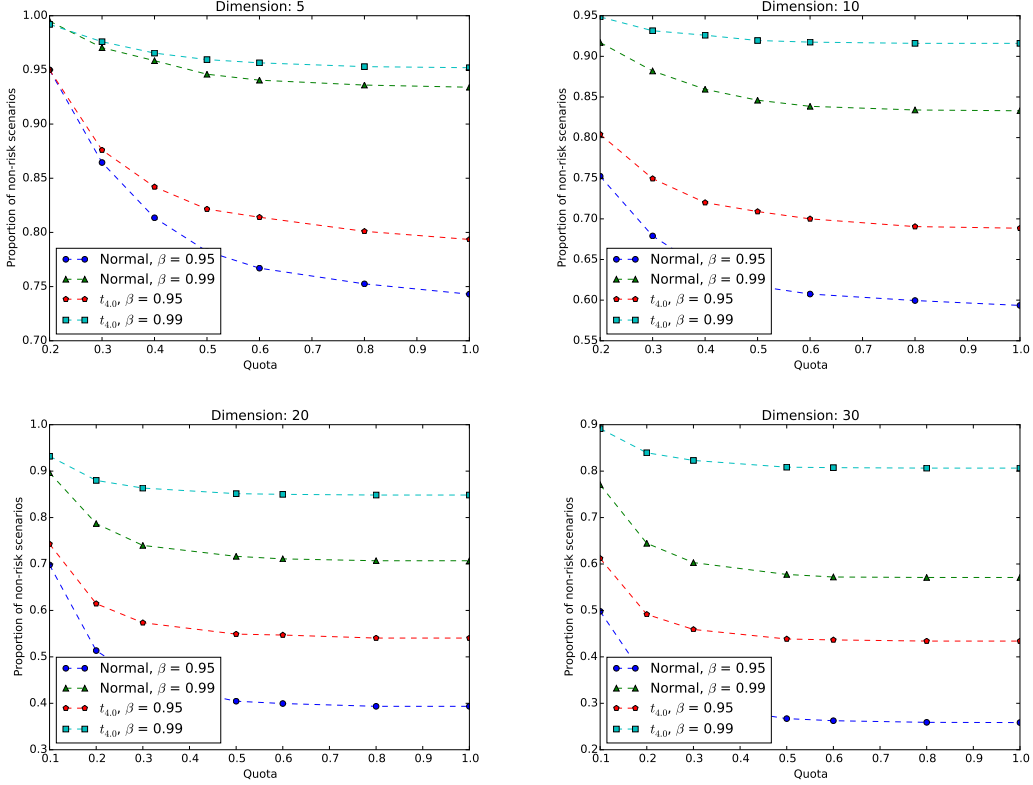


Figure 9: Proportions of non-risk scenarios

results can be found in Appendix A. The first important observation from these is that the proportion of scenarios in the non-risk region, as compared to the uncorrelated case in Figure 6, is surprisingly high; even for $\beta = 0.95$ and dimension 30, this proportion is non-negligible. As expected, the proportion of scenarios in the non-risk region increases as we tighten our quota. However, for higher dimensions the quotas need to be a lot tighter to make a significant difference. The plots also provide further evidence that the t distribution has non-risk regions with higher probabilities than the lighter-tailed Normal distribution. In Figure 9, the non-risk region for the t -distributions has probability around 0.05 to 0.1 higher for dimensions 5 and 10, and around 0.1 to 0.2 higher for dimensions 20 and 30.

6.3 Aggregation sampling

In this section we compare the quality of solutions yielded by sampling and aggregation sampling by observing the optimality gaps of the solutions that these scenario generation methods yield. For this, we use the following version of the portfolio selection problem.

$$\begin{aligned}
& \underset{x \geq 0}{\text{minimize}} \quad \beta\text{-CVaR}(-x^T Y) \\
& \text{such that } x^T \mu \geq \tau, \\
& \sum_{i=1}^d x_i = 1, \\
& 0 \leq x \leq u,
\end{aligned}$$

where μ is the mean of the input distribution (rather than scenario set), τ is the target return and u is a vector of asset quotas. The primary reason for using this formulation over those in (P2) and (P3) is that given a distribution of asset returns it is easy to select an appropriate expected target return τ . For

simplicity, in our tests we set $\tau = \frac{1}{n} \sum_{i=1}^n \mu_i$ which ensures that the constraint is feasible but not trivially satisfied. Notice that in the above formulation we use the deterministic constraint, $x^T \mu \geq \tau$ rather than $\mathbb{E}[x^T Y] \geq \tau$. This is because the latter constraint depends on the scenario set. Therefore, the solution from a scenario-based approximation might be infeasible with respect to the original problem, which makes measuring solution quality problematic.

In this experiment, we test the performance of the aggregation sampling algorithm for three families of distributions: the Normal distribution, the t-distribution and the skew-t distribution.

For each distribution and problem dimension we run five trials using our constructed distributions (as described in Section 6.1). Each trial consists of generating 50 scenario sets via sampling and aggregation sampling, solving the corresponding scenario-based problem for each of these sets, and calculating the optimality gap for each solution which is yielded. For each scenario generation method we then calculate the mean and standard deviation (S.D.) of the optimality gap. For the skew-t distributions, although we are able to evaluate the objective function value for any candidate solution, to find the true optimal solution value (or one close to it), we resort to solving the problem for a very large sampled set of size 100000.

The full results for this experiment can be found in Appendix B. In Figure 10 we have plotted for one trial the raw results for dimensions 10 and 30. We observe that there is a consistent improvement in solution quality in using aggregation sampling over basic sampling. In addition the solution values are much more stable. The improvement in solution quality and stability is particularly big for the t -distributions. This is because the probability of the non-risk region is greater for heavier-tailed distributions as observed in Section 5.

Aggregation sampling also generally leads to better solutions for the skew-t distributions where we are approximating the risk region with a risk region for a t-distribution. However, in three of the twenty constructed skew-t distributions, the solutions are worse. This underlines the point made in Section 4.2 that surrogate risk regions should be constructed carefully.

6.4 Aggregation reduction

The aim of these tests is to quantify the error induced through the use of aggregation reduction. In particular, we calculate the error induced in the optimal solution value. For a given scenario set, we aggregate the non-risk scenarios, solve the problem with respect to this reduced set, and calculate the optimality gap of this solution with respect to the original scenario set.

For these tests we use the same problem as in Section 6.3 and run tests for Normal, t and moment matching distributions. As explained in Section 4, we use the risk region of a Normal distribution to approximate the risk region for moment-matched scenario sets. For each family of distributions and problem dimension we again run five trials for different instances of the distribution. In each trial for different initial scenario set sizes, $n = 100, 200, 500$, and two different levels of tail risk measure $\beta = 0.95, 0.99$, we calculate the reduction error for 30 different scenario sets and report the mean error.

The full results can be found in Appendix C. These show that the reduction error is generally very small, in fact for almost all problems using $\beta = 0.95$, there is no error induced. For $\beta = 0.99$, and the smallest scenario set size $n = 100$, there is a small amount of error (< 0.01) for the Normal distributions, slightly larger errors for the heavier-tailed t distribution (< 0.02), and the largest errors (0.1-0.5) occur for reduced moment-matching scenario sets whose risk regions have been approximated with that of a Normal distribution. However, as the scenario set size is increased, all errors are reduced, and for the largest scenario set size $n = 500$, there is no error induced for almost all problems.

Comparing the reduction errors with the corresponding non-risk region probabilities in Appendix A, we see that the larger errors generally occur for the higher dimensional distributions whose non-risk region has a larger probability. This is to be expected as the larger the non-risk region the more scenarios that are aggregated. In Table 37 in Appendix C we have also included the proportions of reduced scenarios for moment matching scenario sets for which we approximated the risk region with that of a Normal distribution. The proportions of reduced scenarios in this case are generally slightly higher than that of the corresponding Normal distributions. This might suggest that the surrogates for the risk region are slightly too small, but this could equally be explained by the fact that moment matching scenario

sets generally have heavier tails than the corresponding Normal distribution, which, as we observed in Section 5, also leads to non-risk regions of higher probabilities. In either case, the larger errors which are induced by reducing small moment-matching scenario sets could be explained by these increased probabilities.

7 Case study

In this section we demonstrate how our methodology can be applied to difficult problems which may occur in practice. For this, we use problems which are high-dimensional, have non-elliptical return distributions, and use integer variables. Note that the use of integer variables precludes the use of interior point algorithms to solve this problem. For a fixed computational budget we will compare the performance of sampling and aggregation sampling through estimation of the optimality gap. We also demonstrate how ghost constraints can be used to improve the quality of solutions while highlighting the possible pitfalls of this approach.

In these tests we use the following problem:

$$\begin{aligned}
& \underset{x, z}{\text{minimize}} \quad \beta\text{-CVaR}(-x^T Y) \\
& \text{such that } x^T \mu \geq \tau, \\
& \quad x_i \leq z_i \text{ for each } i = 1, \dots, d, \\
& \quad \sum_{i=1}^d x_i = 1, \\
& \quad \sum_{i=1}^d z_i = M, \\
& \quad 0 \leq x \leq u, \\
& \quad z_i \in \{0, 1\} \text{ for each } i = 1, \dots, d.
\end{aligned}$$

This problem is similar to that used in Section 6.3 except that we now use integer variables to bound the number of assets in which we may invest. The extra constraints involving integer variables may change the conic hull of feasible solutions, however the method presented in Section 3.1 for calculating conic hulls of feasible regions cannot handle these. We therefore ignore these constraints when constructing a risk region to use for aggregation sampling. This is acceptable as the resulting conic hull will contain the true conic hull. Again, the random vector Y used in these tests is constructed by fitting Skew- t distributions to return data for companies from the FTSE100 stock index.

In each experiment we find candidate solutions for the above problem by solving large scenario-based approximations: we find one candidate solution for a scenario set constructed by basic sampling, and another for a scenario set constructed by aggregation sampling. The optimality gap for each of these solutions is then estimated by employing the *multiple replication procedure* of [MMW99], which involves solving several n_g (smaller) problems for n_g independent scenario sets constructed by sampling and aggregation sampling as appropriate. Specifically, for $k = 1, \dots, n_g$ denote by Y^k the empirical random vector corresponding to the k -th scenario set, and z_k^* the corresponding optimal solution value. For a candidate solution $\tilde{x} \in \mathbb{R}^d$, and for each scenario set $k = 1, \dots, n_g$ (of size n) the following is a conservative estimate of the optimality gap:

$$G_n^k = \beta\text{-CVaR}(-\tilde{x}^T Y^k) - z_k^*.$$

Now, for $0 < \alpha < 1$ an α confidence interval for the optimality gap is

$$(0, \bar{G}_n + \epsilon_{n_g, \alpha}), \tag{11}$$

where

$$\begin{aligned}\bar{G}_n &= \frac{1}{n_g} \sum_{j=1}^{n_g} G_n^j, \\ S_{n_g}^2 &= \frac{1}{n_g - 1} \sum_{j=1}^{n_g} (G_n^j - \bar{G}_n)^2, \\ \epsilon_{n_g, \alpha} &= t_{n_g-1, \alpha} \frac{S_{n_g}}{\sqrt{n_g}}\end{aligned}$$

and $t_{n_g-1, \alpha}$ is the α -quantile of the (univariate) t -distribution with $n_g - 1$ degrees of freedom. Note that other procedures for estimating the optimality gap exist which only require the solution of one or two problems [BM06], [SB13].

Given the potential dangers in approximating the risk region, and mis-specifying ghost constraints, it is important to verify the quality of a solution by the calculation of its corresponding *out-of-sample* value (out value) [KW07]. That is, we calculate the β -CVaR for our candidate solutions with respect to a large independently sampled scenario set. A bias in our aggregation sampling method may be indicated by a significantly higher out-of-sample value compared to that of sampling. Similarly, the introduction of ghost constraints which are too tight will lead to no improvement in, or a potentially worse out-of-sample value of the new candidate solution. Finally, to aid us in interpreting the results, we include an estimate of the probability of the risk region.

We set our computational budget so that our problems can be solved relatively quickly (a few seconds) on a personal computer. If our problem is of dimension d and we have n scenarios, the number of floating point operations required to calculate β -CVaR for a particular solution is of order $O(d \times n)$. We therefore limit the number of scenarios in our problems so that $d \times n \leq 10000$ to ensure it can be solved quickly. For the estimation of the optimality gap we solve five problems which use half the number of scenarios used to calculate the candidate solution. For both problems we use $\beta = 0.99$, $\tau = 0.01$ and $\alpha = 0.95$. Our aim is to find a solution for which the upper limit of the 95% confidence interval $G_n + \epsilon_{n_g, \alpha}$ on its optimality is less than 0.015.

Case 1: $d = 30$ We begin with a problem of moderate dimension. Using our rules we use a scenario set size of $n = \frac{10000}{30} \approx 3300$ to find our candidate solutions, and for estimating the optimality gap we use a scenario set size of $\frac{n}{2} = \frac{3300}{2} = 1650$. The results are shown in Table 1.

Sampling			Agg. Sampling			
Gap (G_n)	Error ($\epsilon_{n_g, \alpha}$)	Out value	Gap (G_n)	Error ($\epsilon_{n_g, \alpha}$)	Out value	Risk region prob.
0.018	0.005	0.140	0.004	0.002	0.140	0.157

Table 1: Estimated optimality gaps for $n = 30$ with 95% confidence level

The out-of-sample values reveal that the quality of candidate solutions are about the same, however the estimation of the optimality gap using aggregation sampling gives us greater assurance that our solution is near optimal. Using (11) and Table 1, the upper limit of the confidence interval on the optimality gap for aggregation sampling is $0.004 + 0.002 = 0.006$ meets our target of being less than 0.015.

Case 2: $d = 50$ We now increase the dimension of the problem substantially. Our rules for scenario set size now prescribe the use of $n = \frac{10000}{50} = 2000$ for calculating our candidate solutions, and a maximum scenario set size of $\frac{n}{2} = \frac{2000}{2} = 1000$ for estimating the optimality gap. The results are shown in Table 2.

This time both the out-of-sample value and optimality gap yielded by aggregation sampling are smaller than those yielded by basic sampling. Despite this however, the quality of the solutions appear to be much lower than those in Case 1. Using (11) and Table 2, the upper limit of the confidence interval on the optimality gap for aggregation sampling is $0.02 + 0.004 = 0.024$ so does not meet our target.

Sampling			Agg. Sampling			
Gap (\bar{G}_n)	Error ($\epsilon_{n_g, \alpha}$)	Out value	Gap (\bar{G}_n)	Error ($\epsilon_{n_g, \alpha}$)	Out value	Risk region prob.
0.062	0.013	0.191	0.020	0.004	0.164	0.211

Table 2: Estimated optimality gaps for $n = 50$ with 95% confidence level

In an attempt to improve our solution to the previous problem we now add ghost constraints to our problem. As noted in Section 6.2, it is only when constraints become very restrictive that these make a difference to the probability of the risk region and so in the first instance we will add constraints which are tight. We use the constraints $x_i \leq \tilde{x}_i + 0.05$ for $i = 1, \dots, 50$ where \tilde{x} denotes the candidate solution from aggregation sampling for our previous trial. The results for this trial are shown in Table 3.

Sampling			Agg. Sampling			
Gap (\bar{G}_n)	Error ($\epsilon_{n_g, \alpha}$)	Out value	Gap (\bar{G}_n)	Error ($\epsilon_{n_g, \alpha}$)	Out value	Risk region prob.
0.018	0.010	0.168	0.011	0.005	0.165	0.119

Table 3: Estimated optimality gaps for $n = 50$ with 95% confidence level where tight quota constraints have been added to the problem

The proportion of samples in the risk region has roughly halved and so our scenarios are concentrated on a much smaller region of the support. We see this time that the optimality gap of our solutions is much reduced for aggregation sampling, so much so that the error is now within our desired tolerance. However, inspecting the out-of-sample value we see that it has not improved, despite the estimated optimality gap being much lower compared to the previous experiment. In addition, some of the ghost constraints we have added are binding. This strongly indicates that the added constraints may be too tight (they are active), in which case the optimality gaps are not valid with respect to the original problem.

We now try slightly looser ghost constraints: $x_i \leq \tilde{x}_i + 0.1$ where \tilde{x} is still the candidate solution from our first trial without ghost constraints. The results are shown in Table 4.

Sampling			Agg. Sampling			
Gap (\bar{G}_n)	Error ($\epsilon_{n_g, \alpha}$)	Out value	Gap (\bar{G}_n)	Error ($\epsilon_{n_g, \alpha}$)	Out value	Risk region prob.
0.034	0.009	0.169	0.009	0.004	0.162	0.162

Table 4: Estimated optimality gaps for $n = 50$ with 95% confidence level where loose quota constraints have been added to the problem

The out-of-sample value in this trial is significantly improved compared to the previous one and there are fewer binding ghost constraints in our solution. The upper limit of the estimate of optimality gap is now within the desired tolerance. Again, this estimate only applies to the original problem (the problem without ghost constraints) if the ghost constraints are guaranteed to be non-active.

It is generally difficult to guarantee that ghost constraints are non-active, but nevertheless, the example above demonstrates that their inclusion, with careful calibration can lead to significantly improved solutions. In the example above, we only used simple quotas for our ghost constraints and it may be possible to further improve the out-of-sample value by adding lower bounds to investments as well as upper bounds. Additionally, when constructing our ghost constraints we used the same amount of slack for each variable and so one could try varying this for different assets.

8 Conclusions

In the paper [FTW15] we proposed a general approach to scenario generation using risk regions for stochastic programs with tail risk measure. As proof-of-concept we demonstrated how this applied for portfolio selection problems for elliptically distributed returns. In this work, we have presented how this methodology may be used for more realistic portfolio selection problems, and studied under what conditions it is effective.

To find the risk region for our problem, we must be able to describe the conic hull of the feasible region and be able to project points onto this. In the presence of positivity constraints, we were able to describe exactly this conic hull from general linear constraints on our portfolio, and identified that the projection of a point onto a cone requires the solution of a small quadratic program, or a linear complementarity problem. The solution of these small programs becomes a significant bottleneck for high dimensions in our methodology and so one possible avenue of future research would be how one could use second order cones in our risk regions. The projection of a point onto a second order cone is achieved through a simple linear transformation and so can be done much more efficiently than the projection onto a polyhedral cone. The issue to be resolved here is how one can choose a (conservative) second order cone given typical portfolio constraints.

The efficacy of using risk regions for scenario generation depends upon the probability of the risk region: the greater the probability of the non-risk region, the more scenarios that can be aggregated. It follows directly from the definition of risk regions that this probability decreases as the problem becomes more constrained and as the level of the tail risk β increases. In our case study we exploited the former property through the addition of non-binding or ghost constraints to our problem. A more systematic way of selecting ghost constraints, and finding some way to guarantee they are non-active are thus important directions of research. In our numerical experiments we observed that the probability of the risk region decreases for heavier tailed distributions, and in the presence positivity constraints, the probability decreases as the correlations between asset returns increases. It is desirable to develop theory which explain these phenomena.

Finally, we tested the performance of our methodology for solving realistic problems where the return distributions were fitted from real financial return data. Aggregation sampling generally outperformed basic sampling in terms of solution quality and stability. We also showed that aggregation reduction induces almost no error in the solution for reasonably sized scenario sets. These results not only held for elliptical distributions, but also non-elliptical distributions for which we have approximated the risk regions. However, in a small number of cases, the mis-specification of these surrogate risk regions lead to worse results. Thus, research needs to be done to determine how one can choose such more reliable surrogate risk regions for non-elliptical regions.

References

- [AC03] Adelchi Azzalini and Antonella Capitanio. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65:367–389, 2003.
- [ADEH99] P. Artzner, F. Delbaen, J. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.
- [AT02] Carlo Acerbi and Dirk Tasche. On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7):1487–1503, 2002.
- [BM06] Güzin Bayraksan and David P. Morton. Assessing solution quality in stochastic programs. *Mathematical Programming*, 108(2–3):495–514, sep 2006.
- [BV04] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- [Che65] N. K. Chernikova. Algorithm for finding a general formula for the non-negative solutions of a system of linear inequalities. *U.S.S.R. Computational Mathematics and Mathematical Physics*, 5(2):228–233, 1965.
- [CPS92] Richard W Cottle, Jong-Shi Pang, and Richard E Stone. *The linear complementarity problem*, volume 60. Siam, 1992.
- [DR99] R. Dembo and D. Rosen. The practice of portfolio replication. A practical overview of forward and inverse problems. *Annals of Operations Research*, 85:267–284, 1999.

- [FKN89] Kai-Tai Fang, Samuel Kotz, and Kai Wang Ng. *Symmetric Multivariate and Related Distributions (Chapman & Hall/CRC Monographs on Statistics & Applied Probability)*. Chapman and Hall/CRC, 11 1989.
- [FTW15] J. Fairbrother, A. Turner, and S. Wallace. Scenario generation for stochastic programs with tail risk measures. *ArXiv e-prints*, November 2015, 1511.03074.
- [HKW03] Kjetil Høyland, Michal Kaut, and Stein W. Wallace. A heuristic for moment-matching scenario generation. *Computational Optimization and Applications*, 24(2–3):169–185, 2003.
- [Jor96] P. Jorion. *Value at Risk: The New Benchmark for Controlling Market Risk*. Irwin Professional, 1996.
- [JR51] NL Johnson and CA Rogers. The moment problem for unimodal distributions. *The Annals of Mathematical Statistics*, 22(3):433–439, 1951.
- [KBF07] Bahar Kaynar, Ş Ilker Birbil, and J.B.G. Frenk. Application of a general risk management model to portfolio optimization problems with elliptical distributed returns for risk neutral and risk averse decision makers. Technical report, Erasmus Research Institute of Management, 2007.
- [KME00] C.A.J. Klaassen, Ph.J. Mokveld, and B. Van Es. Squared skewness minus kurtosis bounded by 186/125 for unimodal distributions. *Statistics & Probability Letters*, 50(2):131–135, 2000.
- [KW07] Michal Kaut and Stein W. Wallace. Evaluation of scenario-generation methods for stochastic programming. *Pacific Journal of Optimization*, 3(2):257–271, 2007.
- [KW11] Michal Kaut and Stein W. Wallace. Shape-based scenario generation using copulas. *Computational Management Science*, 8(1–2):181–199, 2011.
- [KWVZ07] Michal Kaut, Stein W. Wallace, Hercules Vladimirov, and Stavros Zenios. Stability analysis of portfolio management with conditional value-at-risk. *Quantitative Finance*, 7(4):397–409, 2007.
- [LV92] H Le Verge. A note on Chernikova’s algorithm. Technical Report 635, IRISA, Rennes, France, 1992.
- [Mar52] H.M. Markowitz. Portfolio selection. *Journal of Finance*, 7:77–91, 1952.
- [Mar59] H.M. Markowitz. *Portfolio selection: Efficient diversification of investment*. Yale University Press, New Haven, 1959.
- [MMW99] W.K. Mak, D.P. Morton, and R.K. Wood. Monte Carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, 24:47–56, 1999.
- [RU00] R. Tyrrell Rockafellar and Stan Uryasev. Optimization of conditional value-at-risk. *The Journal of Risk*, 2(3):21–41, 2000.
- [SB13] Rebecca Stockbridge and Güzin Bayraksan. A probability metrics approach for reducing the bias of optimality gap estimators in two-stage stochastic linear programming. *Mathematical Programming*, 142(1–2):107–131, 2013.
- [Tas02] Dirk Tasche. Expected shortfall and beyond. *Journal of Banking & Finance*, 26(7):1519–1533, 2002.
- [You98] Martin R. Young. A minimax portfolio selection rule with linear programming solution. *Management Science*, 44(5):673–683, 1998, <http://pubsonline.informs.org/doi/pdf/10.1287/mnsc.44.5.673>.
- [Zie08] Günter M. Ziegler. *Lectures on Polytopes (Graduate Texts in Mathematics)*. Springer, 4 2008.

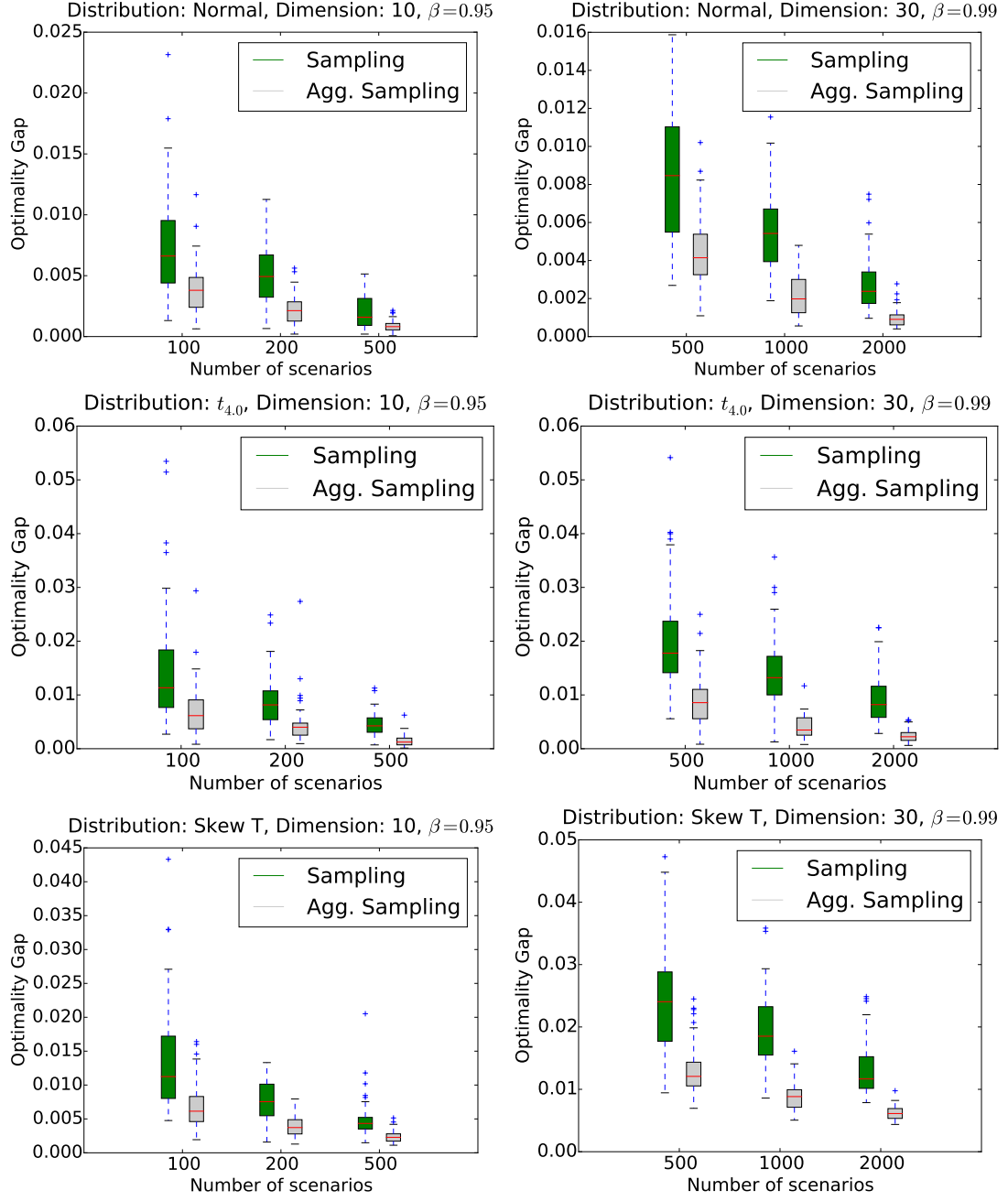


Figure 10: Stability test comparing performance of sampling and aggregation sampling

A Reduction proportion tables

The following tables list the estimated probabilities of the non-risk region for a variety of distributions constructed from real data. See Section 6.2 for details. Each table corresponds to a family of distributions at a given dimension, and each row gives the proportions for a given set of companies. In addition, the distributions corresponding the i -th row of each table of dimension d have been fitted using the same set of companies.

$\beta = 0.95$							$\beta = 0.99$						
$x \leq 1.0$	$x \leq 0.8$	$x \leq 0.6$	$x \leq 0.5$	$x \leq 0.4$	$x \leq 0.3$	$x \leq 0.2$	$x \leq 1.0$	$x \leq 0.8$	$x \leq 0.6$	$x \leq 0.5$	$x \leq 0.4$	$x \leq 0.3$	$x \leq 0.2$
0.743	0.752	0.767	0.782	0.814	0.865	0.950	0.934	0.936	0.941	0.946	0.959	0.971	0.995
0.738	0.744	0.760	0.777	0.809	0.855	0.949	0.922	0.925	0.928	0.934	0.949	0.965	0.992
0.767	0.775	0.793	0.807	0.832	0.872	0.948	0.930	0.932	0.937	0.943	0.953	0.969	0.990
0.763	0.771	0.784	0.801	0.830	0.880	0.951	0.931	0.934	0.944	0.949	0.957	0.973	0.987
0.755	0.763	0.777	0.798	0.829	0.883	0.955	0.927	0.929	0.935	0.940	0.951	0.966	0.991

Table 5: Proportion of reduced scenarios for Normal distributed returns and $d = 5$

$\beta = 0.95$							$\beta = 0.99$						
$x \leq 1.0$	$x \leq 0.8$	$x \leq 0.6$	$x \leq 0.5$	$x \leq 0.4$	$x \leq 0.3$	$x \leq 0.2$	$x \leq 1.0$	$x \leq 0.8$	$x \leq 0.6$	$x \leq 0.5$	$x \leq 0.4$	$x \leq 0.3$	$x \leq 0.2$
0.594	0.600	0.608	0.617	0.637	0.679	0.752	0.833	0.834	0.839	0.846	0.860	0.882	0.917
0.617	0.621	0.632	0.647	0.669	0.703	0.777	0.851	0.852	0.856	0.860	0.868	0.879	0.914
0.506	0.509	0.523	0.534	0.560	0.606	0.689	0.779	0.780	0.787	0.791	0.806	0.837	0.889
0.564	0.566	0.573	0.590	0.615	0.658	0.748	0.827	0.828	0.835	0.846	0.857	0.877	0.921
0.537	0.540	0.552	0.566	0.586	0.624	0.727	0.820	0.822	0.825	0.832	0.843	0.870	0.912

Table 6: Proportion of reduced scenarios for Normal distributed returns and $d = 10$

$\beta = 0.95$							$\beta = 0.99$						
$x \leq 1.0$	$x \leq 0.8$	$x \leq 0.6$	$x \leq 0.5$	$x \leq 0.3$	$x \leq 0.2$	$x \leq 0.1$	$x \leq 1.0$	$x \leq 0.8$	$x \leq 0.6$	$x \leq 0.5$	$x \leq 0.3$	$x \leq 0.2$	$x \leq 0.1$
0.394	0.394	0.400	0.405	0.447	0.513	0.698	0.707	0.707	0.711	0.717	0.740	0.787	0.896
0.325	0.326	0.332	0.342	0.392	0.457	0.635	0.653	0.653	0.655	0.662	0.696	0.740	0.851
0.344	0.344	0.348	0.354	0.389	0.460	0.668	0.648	0.648	0.653	0.656	0.683	0.743	0.870
0.384	0.385	0.390	0.401	0.440	0.507	0.708	0.695	0.695	0.698	0.704	0.740	0.782	0.896
0.417	0.418	0.424	0.432	0.479	0.540	0.738	0.727	0.727	0.730	0.735	0.764	0.813	0.906

Table 7: Proportion of reduced scenarios for Normal distributed returns and $d = 20$

$\beta = 0.95$							$\beta = 0.99$						
$x \leq 1.0$	$x \leq 0.8$	$x \leq 0.6$	$x \leq 0.5$	$x \leq 0.3$	$x \leq 0.2$	$x \leq 0.1$	$x \leq 1.0$	$x \leq 0.8$	$x \leq 0.6$	$x \leq 0.5$	$x \leq 0.3$	$x \leq 0.2$	$x \leq 0.1$
0.259	0.259	0.263	0.267	0.297	0.350	0.498	0.571	0.571	0.572	0.578	0.603	0.644	0.770
0.264	0.266	0.269	0.272	0.299	0.347	0.511	0.587	0.587	0.589	0.591	0.616	0.661	0.790
0.282	0.282	0.286	0.291	0.321	0.378	0.533	0.599	0.599	0.602	0.607	0.631	0.681	0.785
0.247	0.247	0.251	0.257	0.281	0.333	0.502	0.555	0.555	0.556	0.558	0.586	0.630	0.769
0.293	0.293	0.296	0.301	0.324	0.374	0.548	0.583	0.583	0.584	0.587	0.613	0.665	0.802

Table 8: Proportion of reduced scenarios for Normal distributed returns and $d = 30$

$\beta = 0.95$							$\beta = 0.99$						
$x \leq 1.0$	$x \leq 0.8$	$x \leq 0.6$	$x \leq 0.5$	$x \leq 0.4$	$x \leq 0.3$	$x \leq 0.2$	$x \leq 1.0$	$x \leq 0.8$	$x \leq 0.6$	$x \leq 0.5$	$x \leq 0.4$	$x \leq 0.3$	$x \leq 0.2$
0.793	0.801	0.814	0.822	0.842	0.876	0.950	0.952	0.953	0.957	0.960	0.966	0.976	0.992
0.775	0.782	0.796	0.812	0.837	0.877	0.946	0.949	0.950	0.954	0.956	0.961	0.972	0.988
0.808	0.815	0.829	0.841	0.859	0.898	0.953	0.958	0.960	0.962	0.964	0.969	0.980	0.992
0.799	0.808	0.819	0.828	0.855	0.882	0.950	0.949	0.951	0.954	0.957	0.965	0.977	0.990
0.793	0.799	0.809	0.822	0.848	0.887	0.951	0.960	0.960	0.963	0.965	0.969	0.976	0.991

Table 9: Proportion of reduced scenarios for $t_{4,0}$ distributed returns and $d = 5$

$\beta = 0.95$							$\beta = 0.99$						
$x \leq 1.0$	$x \leq 0.8$	$x \leq 0.6$	$x \leq 0.5$	$x \leq 0.4$	$x \leq 0.3$	$x \leq 0.2$	$x \leq 1.0$	$x \leq 0.8$	$x \leq 0.6$	$x \leq 0.5$	$x \leq 0.4$	$x \leq 0.3$	$x \leq 0.2$
0.689	0.691	0.700	0.709	0.720	0.750	0.804	0.916	0.916	0.917	0.919	0.926	0.931	0.949
0.711	0.713	0.719	0.730	0.742	0.769	0.829	0.923	0.924	0.925	0.926	0.930	0.940	0.956
0.616	0.617	0.630	0.640	0.656	0.677	0.754	0.895	0.896	0.898	0.900	0.905	0.915	0.935
0.642	0.642	0.647	0.657	0.672	0.703	0.783	0.896	0.896	0.900	0.904	0.913	0.925	0.941
0.652	0.655	0.666	0.675	0.690	0.723	0.785	0.905	0.905	0.907	0.908	0.913	0.924	0.944

Table 10: Proportion of reduced scenarios for $t_{4,0}$ distributed returns and $d = 10$

$\beta = 0.95$							$\beta = 0.99$						
$x \leq 1.0$	$x \leq 0.8$	$x \leq 0.6$	$x \leq 0.5$	$x \leq 0.3$	$x \leq 0.2$	$x \leq 0.1$	$x \leq 1.0$	$x \leq 0.8$	$x \leq 0.6$	$x \leq 0.5$	$x \leq 0.3$	$x \leq 0.2$	$x \leq 0.1$
0.540	0.540	0.547	0.549	0.574	0.615	0.743	0.849	0.849	0.850	0.852	0.864	0.880	0.932
0.461	0.463	0.467	0.475	0.509	0.560	0.703	0.835	0.836	0.840	0.844	0.858	0.870	0.919
0.506	0.507	0.510	0.515	0.551	0.595	0.753	0.839	0.839	0.839	0.840	0.855	0.874	0.931
0.511	0.511	0.514	0.519	0.562	0.612	0.753	0.860	0.860	0.862	0.865	0.876	0.894	0.939
0.567	0.568	0.572	0.576	0.609	0.657	0.797	0.866	0.867	0.867	0.870	0.881	0.901	0.952

Table 11: Proportion of reduced scenarios for $t_{4,0}$ distributed returns and $d = 20$

$\beta = 0.95$							$\beta = 0.99$						
$x \leq 1.0$	$x \leq 0.8$	$x \leq 0.6$	$x \leq 0.5$	$x \leq 0.3$	$x \leq 0.2$	$x \leq 0.1$	$x \leq 1.0$	$x \leq 0.8$	$x \leq 0.6$	$x \leq 0.5$	$x \leq 0.3$	$x \leq 0.2$	$x \leq 0.1$
0.434	0.434	0.436	0.439	0.459	0.491	0.612	0.806	0.806	0.807	0.808	0.823	0.840	0.891
0.466	0.466	0.468	0.469	0.495	0.532	0.649	0.821	0.821	0.823	0.824	0.838	0.853	0.897
0.443	0.443	0.445	0.448	0.474	0.512	0.637	0.821	0.822	0.822	0.824	0.834	0.854	0.898
0.444	0.445	0.448	0.454	0.470	0.513	0.635	0.812	0.813	0.814	0.814	0.823	0.841	0.889
0.417	0.417	0.419	0.421	0.444	0.487	0.617	0.808	0.808	0.810	0.811	0.823	0.844	0.891

Table 12: Proportion of reduced scenarios for $t_{4,0}$ distributed returns and $d = 30$

B Aggregation sampling tables

The following tables list the relative reduction in the mean and standard deviation of optimality gaps for aggregation sampling compared with sampling for a variety of distributions. See Section 6.3 for more details.

n = 100		n = 200		n = 500	
Mean Imp.	S.D. Imp.	Mean Imp.	S.D. Imp.	Mean Imp.	S.D. Imp.
2.747	2.542	3.226	3.321	3.697	2.871
3.905	4.427	3.226	3.323	3.646	4.439
3.803	2.993	4.889	3.538	4.567	3.927
3.376	3.040	3.402	2.517	5.182	4.357
3.240	3.257	3.432	2.246	4.807	4.708

Table 13: Comparison for $d = 5$, $\beta = 0.95$, and Normal returns

n = 100		n = 200		n = 500	
Mean Imp.	S.D. Imp.	Mean Imp.	S.D. Imp.	Mean Imp.	S.D. Imp.
1.989	1.876	2.670	2.422	2.460	2.495
2.018	2.494	2.711	2.227	3.126	2.864
1.559	1.652	1.736	1.230	2.727	2.678
1.869	2.089	2.275	2.181	2.551	2.731
1.996	2.085	2.285	2.061	2.466	2.828

Table 14: Comparison for $d = 10$, $\beta = 0.95$, and Normal returns

n = 500		n = 1000		n = 2000	
Mean Imp.	S.D. Imp.	Mean Imp.	S.D. Imp.	Mean Imp.	S.D. Imp.
2.357	2.124	2.890	3.039	3.026	2.809
2.504	3.054	2.750	2.839	2.873	2.689
2.308	1.963	2.546	2.854	2.803	2.791
2.341	2.699	2.948	3.369	2.592	2.367
2.802	2.657	3.421	2.494	3.725	3.547

Table 15: Comparison for $d = 20$, $\beta = 0.99$, and Normal returns

n = 500		n = 1000		n = 2000	
Mean Imp.	S.D. Imp.	Mean Imp.	S.D. Imp.	Mean Imp.	S.D. Imp.
1.943	1.842	2.161	2.148	2.901	2.846
1.779	2.195	2.197	2.067	2.590	2.483
1.990	2.227	2.246	2.033	2.405	2.514
2.019	2.012	2.076	2.057	2.010	1.891
1.866	1.769	2.457	1.921	2.853	3.138

Table 16: Comparison for $d = 30$, $\beta = 0.99$, and Normal returns

n = 100		n = 200		n = 500	
Mean Imp.	S.D. Imp.	Mean Imp.	S.D. Imp.	Mean Imp.	S.D. Imp.
2.857	2.661	2.762	1.981	3.500	3.709
3.407	3.431	3.692	3.416	5.572	6.167
4.335	3.062	3.872	4.195	3.244	3.149
4.280	3.748	4.636	6.732	4.974	6.593
2.578	1.773	3.664	3.500	4.019	4.160

Table 17: Comparison for $d = 5$, $\beta = 0.95$, and $t_{4,0}$ returns

n = 100		n = 200		n = 500	
Mean Imp.	S.D. Imp.	Mean Imp.	S.D. Imp.	Mean Imp.	S.D. Imp.
1.899	2.091	2.169	1.805	2.939	2.599
2.078	1.910	2.358	2.229	2.982	2.340
1.996	2.923	2.639	3.126	2.088	1.727
2.658	2.958	2.436	2.222	2.357	2.312
2.080	2.171	1.980	1.232	2.957	2.114

Table 18: Comparison for $d = 10$, $\beta = 0.95$, and $t_{4,0}$ returns

n = 500		n = 1000		n = 2000	
Mean Imp.	S.D. Imp.	Mean Imp.	S.D. Imp.	Mean Imp.	S.D. Imp.
4.142	5.028	4.215	4.383	5.571	5.221
3.039	3.843	4.096	4.346	4.857	6.084
3.378	3.831	4.020	4.267	5.007	5.617
3.722	4.886	3.744	3.247	4.339	5.336
3.616	3.524	4.999	3.739	5.116	6.277

Table 19: Comparison for $d = 20$, $\beta = 0.99$, and $t_{4,0}$ returns

n = 500		n = 1000		n = 2000	
Mean Imp.	S.D. Imp.	Mean Imp.	S.D. Imp.	Mean Imp.	S.D. Imp.
3.035	3.068	2.950	2.547	3.741	4.042
2.359	1.983	3.513	5.068	3.384	3.029
3.507	4.356	2.977	3.966	3.686	4.915
2.950	3.005	3.079	1.964	3.936	4.240
2.228	2.043	3.549	3.227	3.950	4.267

Table 20: Comparison for $d = 30$, $\beta = 0.99$, and $t_{4,0}$ returns

n = 100		n = 200		n = 500	
Mean Imp.	S.D. Imp.	Mean Imp.	S.D. Imp.	Mean Imp.	S.D. Imp.
1.839	1.302	2.508	2.336	2.214	2.177
2.464	3.081	2.043	1.836	2.307	3.333
2.520	2.234	2.744	2.573	-354.685	4.274
2.819	2.009	40.908	2.596	-0.717	3.148
3.510	4.406	3.435	2.886	2.442	2.688

Table 21: Comparison for $d = 5$, $\beta = 0.95$, and Skew T returns

n = 100		n = 200		n = 500	
Mean Imp.	S.D. Imp.	Mean Imp.	S.D. Imp.	Mean Imp.	S.D. Imp.
1.816	1.633	2.040	1.961	2.601	2.441
1.743	2.185	1.945	1.529	1.916	1.888
1.607	1.715	1.700	1.610	1.704	1.713
1.828	2.386	1.846	2.367	1.792	1.843
1.943	2.329	1.976	1.761	2.059	3.480

Table 22: Comparison for $d = 10$, $\beta = 0.95$, and Skew T returns

n = 500		n = 1000		n = 2000	
Mean Imp.	S.D. Imp.	Mean Imp.	S.D. Imp.	Mean Imp.	S.D. Imp.
4.671	4.881	5.379	4.650	8.122	8.535
3.840	5.139	4.029	8.577	3.692	6.850
2.054	1.876	2.160	2.651	2.114	3.654
3.345	2.965	6.542	3.858	22.103	4.560
2.837	3.549	3.330	3.462	3.015	4.290

Table 23: Comparison for $d = 20$, $\beta = 0.99$, and Skew T returns

n = 500		n = 1000		n = 2000	
Mean Imp.	S.D. Imp.	Mean Imp.	S.D. Imp.	Mean Imp.	S.D. Imp.
5.163	2.619	-49.057	3.138	-2.620	2.417
2.092	2.927	2.959	2.445	3.307	3.227
2.676	3.025	3.693	3.686	3.798	5.363
2.443	3.111	2.871	4.439	2.691	2.989
1.845	1.999	2.221	2.684	2.137	4.229

Table 24: Comparison for $d = 30$, $\beta = 0.99$, and Skew T returns

C Reduction error tables

The following tables list the mean error induced by aggregating scenarios in the non-risk region for a variety of distributions. See Section 6.4 for details.

$\beta = 0.95$			$\beta = 0.99$		
$n = 100$	$n = 200$	$n = 500$	$n = 100$	$n = 200$	$n = 500$
0.000	0.000	0.000	0.008	0.002	0.000
0.000	0.000	0.000	0.009	0.001	0.000
0.000	0.000	0.000	0.005	0.002	0.000
0.000	0.000	0.000	0.007	0.001	0.000
0.000	0.000	-0.000	0.007	0.001	0.000

Table 25: Reduction error induced for d=5 Normal returns

$\beta = 0.95$			$\beta = 0.99$		
$n = 100$	$n = 200$	$n = 500$	$n = 100$	$n = 200$	$n = 500$
0.000	0.000	-0.000	0.003	0.000	0.000
0.000	-0.000	-0.000	0.002	0.000	0.000
0.000	-0.000	-0.000	0.002	0.000	0.000
0.000	0.000	0.000	0.002	0.000	-0.000
0.000	0.000	0.000	0.003	0.000	0.000

Table 27: Reduction error induced for d=20 Normal returns

$\beta = 0.95$			$\beta = 0.99$		
$n = 100$	$n = 200$	$n = 500$	$n = 100$	$n = 200$	$n = 500$
0.001	0.000	0.000	0.014	0.005	0.000
0.000	0.000	0.000	0.012	0.003	0.000
0.000	0.000	0.000	0.017	0.002	0.001
0.000	0.000	0.000	0.015	0.005	0.000
0.001	0.000	0.000	0.015	0.004	0.001

Table 29: Reduction error induced for d=5 $t_{4,0}$ returns

$\beta = 0.95$			$\beta = 0.99$		
$n = 100$	$n = 200$	$n = 500$	$n = 100$	$n = 200$	$n = 500$
0.000	0.000	-0.000	0.013	0.003	0.000
0.000	0.000	-0.000	0.017	0.000	0.000
0.000	0.000	-0.000	0.016	0.003	0.000
0.000	0.000	0.000	0.012	0.002	0.000
0.000	-0.000	0.000	0.016	0.002	0.000

Table 31: Reduction error induced for d=20 $t_{4,0}$ returns

$\beta = 0.95$			$\beta = 0.99$		
$n = 100$	$n = 200$	$n = 500$	$n = 100$	$n = 200$	$n = 500$
0.000	0.000	0.000	0.006	0.000	0.000
0.000	0.000	0.000	0.006	0.001	0.000
0.000	-0.000	0.000	0.006	0.001	0.000
0.000	0.000	-0.000	0.004	0.000	0.000
0.000	-0.000	-0.000	0.005	0.000	0.000

Table 26: Reduction error induced for d=10 Normal returns

$\beta = 0.95$			$\beta = 0.99$		
$n = 100$	$n = 200$	$n = 500$	$n = 100$	$n = 200$	$n = 500$
0.000	-0.000	0.000	0.001	0.000	0.000
0.000	-0.000	0.000	0.002	0.000	0.000
-0.000	0.000	-0.000	0.002	0.000	-0.000
-0.000	-0.000	0.000	0.001	0.000	-0.000
0.000	0.000	0.000	0.001	0.000	0.000

Table 28: Reduction error induced for d=30 Normal returns

$\beta = 0.95$			$\beta = 0.99$		
$n = 100$	$n = 200$	$n = 500$	$n = 100$	$n = 200$	$n = 500$
0.000	0.000	0.000	0.015	0.003	0.000
0.000	0.000	0.000	0.015	0.004	0.000
0.000	-0.000	0.000	0.012	0.002	-0.000
0.000	0.000	-0.000	0.017	0.004	0.000
0.000	-0.000	-0.000	0.020	0.003	0.000

Table 30: Reduction error induced for d=10 $t_{4,0}$ returns

$\beta = 0.95$			$\beta = 0.99$		
$n = 100$	$n = 200$	$n = 500$	$n = 100$	$n = 200$	$n = 500$
0.000	-0.000	0.000	0.015	0.001	0.000
0.000	0.000	0.000	0.015	0.004	0.000
0.000	-0.000	0.000	0.013	0.002	0.000
0.000	-0.000	-0.000	0.015	0.004	0.000
0.000	0.000	0.000	0.016	0.004	0.000

Table 32: Reduction error induced for d=30 $t_{4,0}$ returns

$\beta = 0.95$			$\beta = 0.99$		
$n = 100$	$n = 200$	$n = 500$	$n = 100$	$n = 200$	$n = 500$
0.000	-0.000	0.000	0.001	0.000	-0.000
0.000	0.000	0.000	0.002	0.000	0.000
0.000	0.000	0.000	0.000	-0.000	0.000
-0.000	0.000	0.000	0.000	0.000	-0.000
0.000	0.000	0.000	0.001	0.001	0.000

Table 33: Reduction error induced for d=5 Moment Matching returns

$\beta = 0.95$			$\beta = 0.99$		
$n = 100$	$n = 200$	$n = 500$	$n = 100$	$n = 200$	$n = 500$
0.000	0.000	0.000	0.089	0.000	0.000
-0.000	0.000	0.000	0.407	0.003	0.000
0.000	-0.000	-0.000	0.003	0.001	0.000
0.000	0.000	0.000	0.231	0.001	0.000
0.000	0.000	0.000	0.000	0.000	0.000

Table 35: Reduction error induced for d=20 Moment Matching returns

$\beta = 0.95$			$\beta = 0.99$		
$n = 100$	$n = 200$	$n = 500$	$n = 100$	$n = 200$	$n = 500$
-0.000	-0.000	-0.000	0.003	0.000	0.000
-0.000	-0.000	0.000	0.001	-0.000	-0.000
0.000	0.000	0.000	0.509	0.001	0.001
-0.000	0.000	-0.000	0.003	0.000	0.000
0.000	0.000	0.000	0.002	0.000	0.000

Table 34: Reduction error induced for d=10 Moment Matching returns

$\beta = 0.95$			$\beta = 0.99$		
$n = 100$	$n = 200$	$n = 500$	$n = 100$	$n = 200$	$n = 500$
0.000	0.000	0.000	0.205	0.000	0.000
0.000	0.000	0.000	0.111	0.001	0.000
0.000	0.000	0.000	0.206	0.001	0.000
0.000	0.000	0.000	0.218	0.001	0.000
0.000	0.000	0.000	0.071	0.001	0.000

Table 36: Reduction error induced for d=30 Moment Matching returns

$d = 5$		$d = 10$		$d = 20$		$d = 30$	
$\beta = 0.95$	$\beta = 0.99$	$\beta = 0.95$	$\beta = 0.99$	$\beta = 0.95$	$\beta = 0.99$	$\beta = 0.95$	$\beta = 0.99$
0.786	0.919	0.638	0.840	0.481	0.734	0.380	0.646
0.743	0.900	0.623	0.827	0.477	0.741	0.365	0.647
0.761	0.905	0.660	0.869	0.445	0.729	0.381	0.650
0.770	0.907	0.625	0.847	0.455	0.716	0.366	0.655
0.747	0.917	0.640	0.860	0.446	0.712	0.333	0.616

Table 37: Proportions of scenarios reduced for moment matching scenario sets